DOCUMENT RESUME

ED 054 200 TM 000 764

AUTHOR	McNeil, John D.
TITLE	Performance Tests: Assessing Teachers of Reading.
PUB DATE	Apr 71
NOTE	10p.; Paper presented at the Annual Meeting of the California Educational Research Association, San Diego, California, April 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Achievement, *Effective Teaching, *Performance Tests, Post Testing, Pretesting, *Reading, Reading Skills, *Teacher Evaluation, Teacher Influence, Teaching Skills, *Testing, Test Reliability, Test Validity

ABSTRACT

The rationale and description of tests of teaching power by which teachers have an equal chance to show their relative ability to effect pupils' achievement in reading skills are discussed. Illustrations of these performance tests and a means for administering them - "teaching faires" are also presented. Data in support of teaching performance tests in reading are reported along with information about teachers resistance to such tests. Suggestions are made for further work with tests of teaching power. (Author/AG)

ED 054200

PERFORMANCE TESTS: ASSESSING TEACHERS OF READING

John D. McNeil
University of California, Los Angeles*

INTRODUCTION

We believe that the teacher makes a difference in the child's progress in learning to read. Not all teachers are equally able to help pupils both achieve mastery of critical skills and develop positive attitudes toward reading. Unfortunately, administrators who are charged with responsibility for selecting and improving teachers seldom have had access to that information that enables one to recognize those teachers who are indeed superior in teaching. Courses taken, grades, age, experience, personality as judged through interview, observation, and supervisor's ratings--these and similar kinds of information have been found inadequate for making judgments about which of several teachers is the most competent.(Morsh & Wilder, 1953). Admittedly, pupil growth is the ultimate criterion for assessing teacher effectiveness. However, it is unsound to rank teachers on this criterion when they have not been confronted with a comparable set of teaching conditions including factors such as common instructional tasks, teachable children, and time allowed for teaching. The problem, therefore, is to design tests of teaching power by which teachers have an equal chance to show their relative ability.

*Special recognition is due V. Downs and U.C.L.A Coordinators of Student Teaching (elementary) who participated in the design and conduct of "Teaching Fairs," an original contribution to performance testing.

TM 000 164

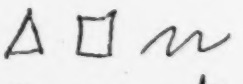
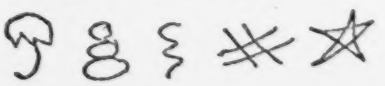
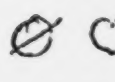
U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

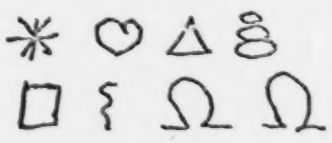
GENERAL DESCRIPTION OF PERFORMANCE TESTING

Performance tests are one answer to the problem of identifying the effective instructor. Typically, performance testing means giving a number of teachers identical instructional tasks (objectives) and a sample of a measure to be administered to pupils after the teaching has occurred. The instructional tactics are left to the teacher. Frequently, the objective is novel to both teacher and pupils, thereby eliminating major "contamination" from previous exposure to the subject matter and aiding in the problem of experimental control. The teachers are allowed a specific period of time for planning the lesson(s) and for the teaching. Groups of learners are assigned to the teachers as pupils. These learners are drawn from a common population and are randomly assigned to a group for instruction. Following the instructional period, all pupils are assembled to complete a test which measures pupil attainment of the instructional objectives. The mean of the test scores earned by pupils taught by a given teacher indicates that teacher's standing in ability to teach the predetermined skill or concept.

EXAMPLE OF PERFORMANCE TEST

(Task one: A test to provide evidence of a teacher's ability to help pupils break a code. Although the "letters" in this teaching exercise are artificial, the task is not altogether unlike that of recognizing long and short vowel sounds in printed words that follow consonant-vowel-consonant and consonant-vowel patterns. No assumption

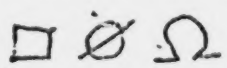
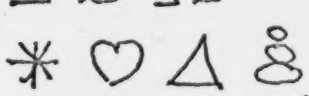
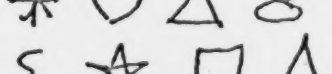
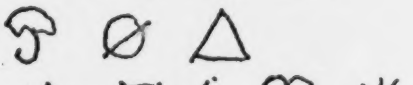
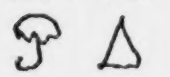
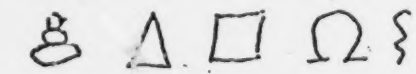
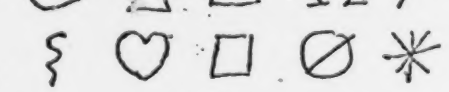
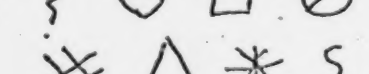
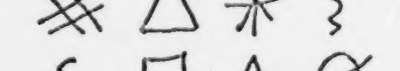
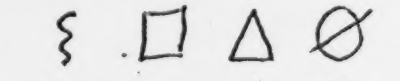
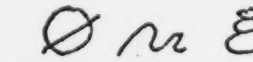
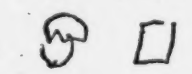
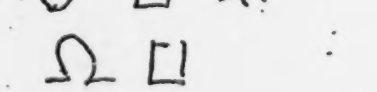
is made however that children who master the contrived task will be able to perform with conventional letters.) Teaching time 15 minutes. Objective: Given a list of written words in code, the pupil will be able to circle those which contain a short bling sound produced by GBG pattern. In this exercise, a short bling sound occurs only when the bling is both immediately preceded and immediately followed by a glonk. The following are symbols standing for blings:  The following are symbols standing for glonks:   All words are made up of glonks and blings. Sample test item: Circle the words which have a short bling sound.



Post Test: (To be given to pupils. Not available to teachers)

The tester will read the directions aloud.

Directions: Circle each word below which has a short bling sound.

1. 
2. 
3. 
4.
5. 
6.
7. 
8. 
9. 
10. 
11. 
12. 
13. 
14. 
15. 

Circle the answer that tells how you feel about the questions:

A. Do you want more lessons from the teacher who taught you the code? Yes No

B. Do you want more lessons like this code lesson? Yes No

In addition to the test presented above, three other teaching tests have been used with more than 200 teachers in training. These tests require the teacher to teach tasks analogous to:

(a) recognizing new words and then selecting the one word from among several that will best complete a sentence composed of these words, a task demanding both skill in word recognition and ability to apply structure of language in completing sentences. (Test 2)

(b) indicating the sound value of a given letter when there is a single letter c with two sound values--/c/ and /k/(c as in cent; c as in cat). (Test 3)

(c) determing pronunciation of initial vowels in words by using the "silent e rule." (Test 4)

Each of the tests was developed in accordance with the following guidelines:

1. The objective or task should be analogous to an important skill in word recognition (validity).

2. The task should require the learner to apply his learning to fresh instances (no teaching to the test).

3. Evidence should be collected indicating the child's attitude or predisposition toward both teacher and task.

4. The task should be complex enough to allow teachers to make decisions regarding such matters as reinforcement, pacing, relevant and irrelevant practice, identification of prerequisite skills, and sequencing.

5. The task should be one that pupils with competent help can master within the time allowed, yet must be difficult enough so that it will discriminate among teachers.

TEST TRYOUTS

Three populations of teachers have served as subjects: those enrolled in a methods course in reading, those completing an initial assignment as student teachers, and those finishing a second teaching assignment. These teachers were grouped--approximately 20 to a group--and directed to a school for participation in a "Teaching Fair." A Teaching Fair is akin to traditional fairs where skilled persons enter competitive contests, publically displaying their expertise. The fairs took place in schools ranging from inner city schools where pupil performance on standardized tests was among the nation's lowest to wealthy suburban schools where reading achievement scores ranked in the top tenth percentile on national norms. Schools, pupils, and task (in its analogous form) were unfamiliar to the teacher. About one hundred twenty children at each site from second, third and fourth grades served as the learners. Children identified as having exceptional intelligence, emotional behavior, and language backgrounds were not included in the population taught. All teachers taught at the same time, the teachers usually teaching in a common location such as a lunch area. Each teacher taught first a group of three children (randomly drawn from the pupil population) and then after a fifteen minute recess taught a second group of three. Mass testing of pupils followed immediately after the lesson and was conducted by independent auditors. Pupil responses were corrected and the teacher's total score was compiled

for both correct items and positive responses toward teacher and task. Teachers were then ranked within their groups. The results were available for those hiring teachers in the case of student teachers seeking employment and for grading purposes in the methods course.

WHAT HAVE WE LEARNED?

GENERALIZATIONS

A caveat is in order. The particular tests described in this paper do not represent all alternatives possible under the rubric performance testing. Illustrations of other formats can be seen in the works of Justiz (1969), Popham (1971), and Taneman (1970). Test developers have many other options such as increasing the number of days required for teaching and providing, in addition to objectives and sample test items, instructional resources and distractors. With respect to the present tests, there are data indicating test validity, reliability and practicality.

1. Validity. The tests are drawn from reading skills generally recognized as important in learning to read. (See consensus of reading skills as determined by Otto and Peterson, (1969). It is true, however, that the tasks are analogous to the reading skills and not identical to them. It is assumed that the teacher who can succeed in communicating the key to breaking the artificial code can also communicate the key for breaking the conventional code.

2. Reliability. First session scores of fifteen teachers on Test 1 above, were compared with these teacher's scores from a second session. The correlation between the scores was .521, significant at the .05 level. Also, correlations between teachers' performances on the different tests given ten weeks apart with different kinds of pupils were positive. For instance, thirty teachers took Test 3 at the end of their methods course and completed Test 4 ten weeks later after a student teaching assignment. Their scores showed a Pearsonian r of .388 $p(<.05)$. As indicated in Table 1 below, one could have made a probable prediction about the likelihood of high achieving and low achieving teachers (top 25 percent and bottom 25 percent) making a similar showing on a second test weeks later. The chances of these teachers maintaining their level were greater than three to one.

Table 1

Chi Squares for High and Low Teacher Performance on Tests of Ability to Teach Two Different Tasks of Reading

TASK 3		TASK 4 High	TASK 4 Low
13 high		9	4
13 low	$X^2 = 4.54$ $p(<.05)$	4	9

TASK 2		TASK 4 High	TASK 4 Low
16 high		12	4
16 low	$X^2 = 4.40$ $p(<.05)$	6	10

3. Utility. There are three ways in which utility has been shown. First, employers have stated that in making a decision about which of several teachers to hire, the information about a teacher's performance on the test relative to his peers is of value along with other kinds of information. Second, some teachers who did not obtain satisfactory results the first time they took such a test have been able to study the demands of these tests and to analyze their own practice with respect to these demands, thereby improving in their ability to perform. Third, the tests have been used as a research tool. Taped records of the teaching carried out by high and low scoring teachers have been analyzed and promising instructional procedures have been identified. These procedures are now being systematically manipulated to verify their importance.

OPPOSITION TO USE

Any information which might be used to assess a teacher and for making decisions about his employability is likely to generate anxiety. Consequent resistence to the performance test as an information gathering scheme seems to take these forms:

1. A few teachers reflect their "equalitarian bag" and minimize aptitude. They play down the fact that not all teachers are equally competent to serve children after a fixed period of training. There are those who state that it is the business of the training program to ensure that all teachers succeed.

2. There are teacher who want to be judged solely on subjective criteria or on the basis of their efforts, not results produced. They feel more confident in competing with their peers on the basis of personality and hard work. Those who have had a history of success in winning friends and influencing others probably believe they have a better chance of competing for a job on the basis of the general impression they make on supervisors, principals, and interviewers than when forced to compete on the basis of their ability to effect desired changes in learners. Other teachers feel that because they worked hard, even though they accomplished little with children, they are good teachers.

3. Some teachers have claimed that they did not receive equal opportunity to succeed on the test. When it can be verified that indeed their pupils or situations were not representative, these teachers have been given another chance on a different test. Usually teachers begin to question their own performance rather than to blame the pupils for the failure when it is shown that other teachers get successful results with the same group of pupils on a related task under similar conditions.

4. Does not deal with goals that are not measured easily (i.e., do not lend themselves to behavioral objectives) and these goals may be more important.

5. Does not take into account the rapport (or lack of it) between pupil or teacher built up over time.

NEXT STEPS

Further test development is needed. Variations in test construction should be created and tried out. Also, research should be undertaken to find out how generalizable the results of performance tests are; e.g., What is the relation of short fifteen minute performance scores to

semester goals? What is the relation of a teacher's success on classes of reading skills in addition to the relation of his success on reading tasks within a class of skills? Then too, the attitude and role of community groups, teachers' organizations, and personnel commissions with respect to performance tests deserve study. One likely pressure in favor of performance testing is the recent Supreme Court decision barring discriminatory job testing. This action should result in school employers demanding tests that will provide information predictive of or correlated with important elements in the teaching of reading--- the job for which the candidate is being evaluated.

REFERENCES

Justiz, T. A reliable measure of teacher effectiveness. Educational Leadership Research Supplement, October 1969, 3 (1), 49-55.

Morsh, J. E. & Wilder, E. W. Identifying the effective instructor: A review of the quantitative studies 1900-1952. Project No. 7714. Task No. 77243. Chenute Air Force Base, Illinois: Air Force Personnel and Training Research Center, 1953.

Otto, W. & Peterson, J. A statement of skills and objectives for the Wisconsin prototypic system of reading skill development. Working paper no. 23, Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, August 1969.

Popham, W. J. Performance tests of teaching proficiency: Rationale, development, and validation. American Educational Research Journal, January 1971, 8 (1), 105-117.

Taneman, I. A teaching power test: A method of using pupil gain as a criterion in teacher education. Unpublished doctoral dissertation, University of California, Los Angeles, 1970.